DOCUMENT RESUME

ED 373 097 TM 021 988

AUTHOR Dynarski, Mark
TITLE Implementing Rigorous Evaluations of Education Interventions: Findings from Two Federal Demonstration Programs.
PUB DATE Mar 94
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Control Groups; *Demonstration Programs; Dropout Programs; Evaluation Methods; *Evaluation Utilization; *Federal Government; Intervention; Nontraditional Education; *Program Evaluation; *Research Design

ABSTRACT

Since 1988 five evaluations of Federal demonstration programs in education have been implemented that had random-assignment designs to measure program impacts. The implementation of two of these evaluations, the evaluation of the School Dropout Demonstration Assistance Program and that of the Alternative Schools Random Assignment Program, are explored to lend support for several conclusions regarding random assignment designs. The first is that random assignment can be implemented in a variety of settings and at a scale that is suitable for measuring impacts precisely, but that it is generally poorly understood by educators. Consequently, it is difficult to implement without a great deal of discussion and negotiation. A second conclusion is that program staff fear that they will lose control of who is admitted to the program. The mechanics of random assignment need to be tailored to this concern. A final conclusion is that random assignment is most likely to fail when the pool of applicants is inadequate to support creation of a control group. When programs experience shortages of applicants, random assignment is not desirable. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 097

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MARK DYNARSKI

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IMPLEMENTING RIGOROUS EVALUATIONS OF EDUCATION INTERVENTIONS: FINDINGS FROM TWO FEDERAL DEMONSTRATION PROGRAMS

Mark Dynarski
Mathematica Policy Research, Inc.

March 1994

Paper presented at the annual meetings of the American Educational Research Association, New Orleans, LA, April 4-8, 1994. I wish to thank Stuart Kerachsky for his comments. The opinions in this paper do not reflect the views or policies of Mathematica Policy Research, Incorporated.

TM021988

BEST COPY AVAILABLE

IMPLEMENTING RIGOROUS EVALUATIONS OF EDUCATION INTERVENTIONS: FINDINGS FROM TWO FEDERAL DEMONSTRATION PROGRAMS

Mark Dynarski
Mathematica Policy Research, Inc.

Since 1988, five evaluations of federal demonstration programs in education have been implemented with random assignment designs to measure program impacts.[1] This represents a remarkable growth in the use of random assignment for evaluations of education programs. The growth has been accompanied by lessons about using random assignment that may be useful for evaluators and policy makers.

I want to focus my remarks today around three conclusions that emerge from my experience implementing random assignment in two of these recent evaluations, the Evaluation of the School Dropout Demonstration Assistance Program (which I will call the *Dropouts* evaluation), and the Alternative Schools Random Assignment Evaluation (which I will call the *Alternative Schools* evaluation). The three conclusions are

(1) Random assignment can be implemented in a variety of different settings and at a scale that is adequate for measuring impacts precisely. However, random assignment is poorly understood by educators, who are likely to view random assignment negatively without understanding what it is or how flexible it can be. Consequently, efforts to implement random assignment are likely to require a large amount of discussion and negotiation.

(2) In terms of the challenges posed to evaluators who want to implement random assignment, the most important is the concern of program staff that they will lose control over who is admitted to the program. The mechanics of random assignment need to be tailored to address this concern. Ethical concerns about denying services to students are also raised by local program staff, but these can be addressed in a straightforward way and are not likely to block implementation.

(3) Random assignment is most likely to fail when the pool of applicants is inadequate to support creating a control group. By their nature, school districts are particularly unable to "market" special programs to attract applicants, which is especially important to do

[1]The programs are the Carl Perkins Vocational Education demonstration program, Even Start, the Alternative Schools Demonstration Program, the School Dropout Demonstration Assistance Program, and Upward Bound. In addition, recent evaluations of JOBSTART, Career Academies, and Job Corps, programs with substantial education components, have also been evaluated or are being evaluated using random assignment designs. The evaluation of the National Adult Workplace Literacy Program may also use random assignment for some of its participating programs.

when students participate in a program voluntarily. As a result, some programs experience shortages of applicants, which makes random assignment undesirable.

The Context

A brief description of the two programs that are being evaluated will help to set the context for these conclusions. The Alternative Schools program began in 1988 in seven sites--Newark, Detroit, Cincinnati, Denver, Wichita, Stockton, and Los Angeles--with funding from the U.S. Department of Labor. The program provided $800,000 over two years to local school districts to create alternative high schools that were to focus on improving the basic skills of at-risk students in a positive academic setting. Program components were based on High School Redirection in Brooklyn, and included small school settings (about 300 students), small class sizes, abundant counseling and services, and special assistance for students with poor reading skills. Program eligibility criteria included being one or more years behind grade level, poor grades, poor attendance records, or past histories of delinquency or drug use. Currently, all but one of the schools is still operating (the federal grant ended in 1990). The Denver school was closed by the district in 1992 due to local budget pressure.

The Dropouts program began in 1988 and the current round of funding began in 1991, with 65 programs receiving grants from the U.S. Department of Education ranging from $100,000 to $1,500,000 a year for four years. The programs are following two general approaches for addressing the dropout problem. The first approach--termed the *targeted* approach--involves providing services such as instruction, counseling, and social service referrals for a defined population of at-risk students. The second approach--termed the *restructuring* approach--involves school-wide reform for a group of schools generally centering around a high school and its feeder middle and elementary schools. The reforms include changes in instruction, curriculum, governance, and articulation. Of the 65 programs, 57 adopted the targeted approach, with grants averaging about $500,000 a year, and 8 adopted the restructuring approach, with grants averaging about $1,000,000 a year. The programs were funded for four years and are now entering their last year of funding.

The Alternative Schools evaluation involved implementing random assignment in 6 sites (Los Angeles was dropped at an early stage due to its inability to implement the model), with a target of 400 sample members in each site over a two-year intake period. The Dropouts evaluation involved implementing random assignment in 20 targeted sites, also with a sample target of 400 sample members in each site over a two-year intake period. The longitudinal data collection efforts for both evaluations included at least two rounds of follow-up student questionnaires, student records abstraction, and, for the Alternative Schools evaluation, on-site administration of a basic skills test (the Test of Adult Basic Education). Data collection activities are under way for both evaluations and preliminary results will be available within a year.

Implementation Factors

Ultimately, random assignment was implemented successfully in 17 of the 26 sites, if timeliness is considered, in 19 of 26 sites, if timeliness is not considered, and in 19 of 23 sites, if we drop from the base sites whose funding was cut and who are unable to participate in the evaluation as a result.[2] So, depending on how it is counted, the success rate for implementing random assignment in the two evaluations is between 65 and 85 percent. Sample sizes are large: to date, the combined sample for the two evaluations exceeds 7,000. So, clearly, *random assignment can be implemented in the context of education programs.*

However, over the course of the implementation effort, two key observations emerged that are relevant to future efforts to implement random assignment. The first observation is that educators and program staff are ill-disposed towards random assignment in particular and impact evaluation in general. A common approach used by evaluators in arguing for random assignment is to say that "we

[2]For the Dropouts evaluation, sites were required to do a substantial amount of data collection to support the evaluation, and initially they received funding in their grants to support these activities. However, the Department of Education cut grants in the second year of funding, and some programs offset the reduction by using funds slated for data collection activities to support program services instead.

want to know whether the program works, and random assignment is the best tool for the job." This approach is properly scientific in that it adopts the skeptical stance that evidence is needed before a program can be judged to have worked. This approach also presumes that, like evaluators, educators and program staff are skeptical about whether their programs work and want program impacts to be measured in the most accurate manner. I would say the opposite is more true: program staff already believe that their programs work, and consequently they don't see much reason to use random assignment. In fact, from their perspective random assignment might only show that a program does *not* work, which is knowledge that might satisfy researchers but would leave program staff feeling unhappy and threatened.[3]

Faced with program staff who can be made to understand random assignment but who are threatened by it, the best recourse of the evaluator is to use the leverage at their disposal: the threat that failure to comply with random assignment could result in reduced grant funds. In practice, this means that evaluators should emphasize in initial meetings with program staff that the evaluation is required under the conditions of the grant and that through careful discussion, aspects of the evaluation that are particularly bothersome can possibly be modified. There may be some discussion about how the agency is inflexible and demanding, and that the evaluators can't do anything about the overall master plan to implement random assignment but will do what they can to reduce the burden imposed by it.[4] Placing the blame for a predicament on a third party who is not at the table

---

[3]A natural consequence of feeling threatened by impact evaluation is to argue that it is not important to measure impacts. Program staff therefore push for "formative" evaluations, which could possibly show that their programs do not work well in an organizational sense but which generally result in suggested improvements to the program that will take time to implement, or they argue that their program is designed to improve affective outcomes, like self-esteem or attitude, rather than quantitative outcomes, like grades or test scores. These arguments are moot in the evaluations described here, which included thorough process analyses and whose data collection instruments included scales designed to measure affective outcomes. But these aspects of the overall evaluation are worth emphasizing in discussions with program staff.

[4]A principal at an alternative high school once demanded to know from me whose idea it was to use random assignment. I responded that it was policy at the U.S. Department of Labor to use random assignment for all its evaluations. The principal said "At least it wasn't your idea." I think

is a time-honored way to create goodwill in a negotiation. Given the strict hierarchical structure of school districts, it is also most useful to negotiate with the highest-ranking administrators who have authority over a program. If buy-in does not happen among high-ranking administrators, it is unlikely to happen at the level of staff who are operating the program.[5]

The second observation is that educators and program staff generally have no idea how random assignment works in reality, but they have their own view of how it works and they are opposed to doing it that way. In fact, I think the most commonly held perception of random assignment is that it entails selecting a group of students randomly from some population, and then directing them into a special program. Furthermore, most staff probably believe that students are not allowed to leave the program without permission of the evaluators. Understandably, staff are opposed to selecting students for programs in this fashion.

Of course, in practice, random assignment operates only with students who are deemed appropriate for the program as the program naturally operates, and students selected for a program are free to enter or exit the program as they like. Program staff are usually relieved to know this, but because even a modest dropout prevention program may have 10 or more staff, it can take a considerable effort before all staff lose their prejudice toward random assignment.

---

the principal was more comfortable working with me to implement random assignment knowing that I was not the real cause of the problem.

[5]For the Alternative Schools evaluation, a decision was made early in the implementation effort to first approach principals of the alternative schools about conducting random assignment. The strategy was that if the principals bought in to using random assignment, then implementation could proceed. If they did not, then the next highest ranking staff person above the principal would be contacted, until implementation was achieved. The strategy lead to lengthy delays in implementation because principals generally expressed reluctance and eventually higher-ranking staff needed to be involved, which took time. Implementation was smoother in the schools where principals immediately turned the negotiation over to a higher-ranking administrator.

Ethical Concerns Can be Addressed

At the beginning of the implementation effort, the evaluation team members worked under the assumption that program staff would be concerned about the ethical problem of denying services to control group members. Responses to ethical concerns were prepared that highlighted the fact that program slots were scarce and that random assignment was an equitable method for allocating scarce slots.

In actual experience, ethical concerns were frequently raised by program staff, but were likely to recede quickly after evaluators explained the fairness of using random assignment to allocate scarce slots. This may be attributable partly to the leverage strategy described above, in which evaluators said that random assignment had to be done and that discussion should center around how to make it fit the demands of the program. This strategy does not leave much room for long discussions centering on ethical concerns. But it is also consistent with a scenario in which program staff attempt to put up resistance to the evaluation and the ethical problem comes to mind but it is not heartfelt. It is interesting that few program staff perceived that reducing overenrollment by not informing eligible students about a program is a form of allocating scarce slots. When evaluators argued that not informing eligible students about a program could be construed as an unfair allocation, the ethical concerns were deflected back onto the program staff. The ethical issue became a draw and discussion moved on.

A more serious sticking point for implementation is the natural tendency for random assignment to treat all applicants alike. In making determinations about who should be admitted to a program, many programs give differential weight to students who are more seriously at-risk. Generally, programs were more likely to admit seriously at-risk students, though some programs were less likely to admit seriously at-risk students (typically, other programs existed in the local area that were more suitable for these students than the program being evaluated).

To insure that random assignment did not skew the mix of students away from the mix the program wanted to serve, random assignment had to be tailored to the individual programs. In practice, the three most common ways in which random assignment was tailored were to stratify applicants according to criteria imposed by program staff, to use differential random assignment probabilities for particular strata, and to give program directors the flexibility to admit directly a small number of applicants who have special circumstances (known as "wild cards"). Analytic complications are introduced by these accommodations, but the complications are offset by the greater likelihood of implementation.

Applicant Shortages Are a Barrier to Random Assignment

The primary reason that random assignment was not implemented successfully in all programs where it was attempted was that programs overestimated the number of applications they would receive. Random assignment is feasible only where there is a real surplus of applicants. Ideally there should be twice as many *eligible* applicants as the program can hold (this enables a 1 to 1 assignment rate to be used), and no less than 50 percent more than it can hold (this enables a 2 to 1 assignment rate to be used). However, though programs typically believe they will be flooded with applications when they open their doors, the reality can be very different. For example, four of the six programs in the Alternative Schools evaluation had applicant shortages that in some way led to difficulties for random assignment.

Reasons for applicant shortages are not hard to find. Eligible students may not apply because they have never heard about the program, or they have heard negative reports about it, or they think alternative programs are only for stupid kids or troublemakers. Programs also overstate the number of eligible students dramatically when applying for grants, because grant competitions frequently award points for demonstrated need. So, for example, an alternative high school will demonstrate need by calculating the number of dropouts in the local area. This is somewhat like calculating the

demand for toothpaste in a local area by counting up the number of people who have teeth. In fact, other programs compete for applicants, and some eligible students do not want to come back to school, which is where they failed in the first place. When these factors are considered, the real pool of eligible students may be much smaller than demonstrated need.

Competing for applicants may require marketing a program, such as by using posters and fliers, public service announcements, and press releases. These activities can be done with modest budgets. By nature, however, school districts may be uncomfortable marketing their special programs. This may be attributable to the fact that as government agencies, they generally do no marketing of any kind.[6] As a result, school districts can be slow to react to shortages of applicants, and internal tensions in a school district can act as barriers to recruiting more applicants.[7] These are important considerations for evaluators because for the most part, evaluators can do little to solve these problems. In initially discussing random assignment with a school district, evaluators are well-advised to probe extensively to understand *where the applicants are coming from,* and to be skeptical about any claims that there will be "no problems at all" getting applicants. Think of the number of small businesses that have failed because they were over-optimistic in their sales projections.

---

[6]This is not true for private schools, of course, and it is also not true for community organizations, some of whom operate education programs. Not surprisingly, some of the most aggressive outreach efforts we observed were created by community organizations.

[7]For example, one alternative high school in a large urban district faced a persistent shortage of applicants because it relied on staff in comprehensive high schools to refer eligible students to it. However, staff of the comprehensive high schools did not like the fact that the alternative school received a much larger per-student budget allocation, and so they would refer only the most seriously at-risk students to the alternative high school, which the school was reluctant to admit because of their potential to disrupt the school. District administrators were reluctant to market the program publicly for fear of antagonizing principals of the comprehensive high schools. Evaluators do not have much leverage to affect a situation like this.

Conclusions

The considerable power of random assignment designs is now being brought to bear on education programs for at-risk youths. The results of these efforts will no doubt lead to clearer thinking about program design, and ultimately to better programs.

A proven ability to implement random assignment designs for federal programs in education may lead to a greater reliance on them in the future. The observations noted above may help lead to smoother implementation of rigorous evaluations. Briefly, the observations were that (1) the perceptions of random assignment among educators are uniformly negative and much of the work of implementing random assignment involves creating a more positive image of it; (2) the best negotiating strategy with program staff is to use the leverage created by the relationship between evaluation and continued funding, and to argue that the real discussion should be is about how best to do random assignment rather than whether it should be done; and (3) applicant shortages cause serious difficulties for random assignment and evaluators should focus attention early on understanding whether programs can really generate sufficient applicants to create a control group. The good news is that ethical concerns do not seem to present much difficulty.

I remain concerned about the acceptance of random assignment among educators and program staff at the local levels. There is no doubt that the U.S. Department of Education is now committed to using random assignment (the U.S. Department of Labor has been committed to random assignment since the early eighties). However, random assignment can be used by local school districts to a much greater extent than it is. A strong push by ED and DOL to disseminate the findings of their random-assignment evaluations and to link the power and influence of the findings to the use of random assignment designs may help to broaden the use of random assignment by local educators.

Ultimately, however, I think a broader acceptance of random assignment evaluations requires that educators adopt a more skeptical view of programs than they currently have. As long as

educators believe that doing something is sufficient because it is better than doing nothing, the role of rigorous evaluation of education programs will be limited. Instead, educators should be striving to do the best thing, and that is where information from rigorous evaluations is most valuable.